

# **Sistemas de Inteligencia Artificial**

TPE 2 - Redes Neuronales

# Objetivos

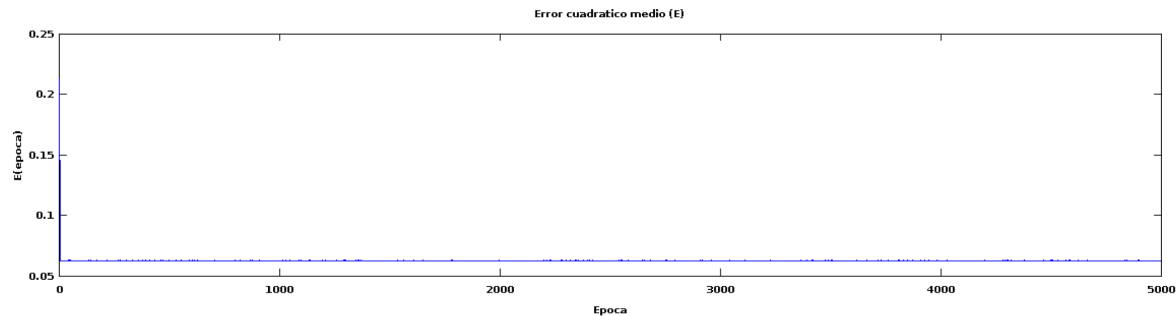
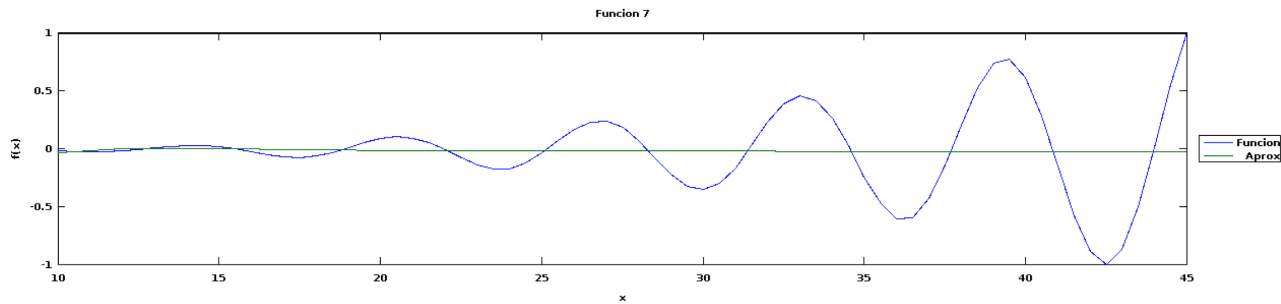
- Crear una red neuronal capaz de estimar el valor de una función  $f$  de dominio e imagen reales
- $f(x) = \sin(x) * x^3 + x/2$ , con  $x \in [10, 45]$
- Implementar y comparar mejoras a la red neuronal:
  - $\eta$  Adaptativo
  - *Momentum*
- Encontrar la arquitectura óptima para el problema.
- Encontrar alguna combinación que aproxime la función con un  $E(W) < 0.0001$

**La entrega anterior no  
funcionaba bien ¿por qué?**

**No se realizaba un *feed forward* de la red antes de realizar el cálculo del error y graficar la función**

**Una vez que se solucionó el problema, se procedió a probar la red. Esto no funcionó...**

# Primera prueba



$$\eta = 0.001$$

$$\beta = 1$$

$$\text{arq} = 1-35-10-1$$

$$g = \tanh$$

$$g \text{ salida} = \text{lineal}$$

$$\text{Épocas} = 5000$$

**¿Por qué no funcionó?**

**Los patrones eran muy grandes lo que causaba que la suma pesada de los pesos por los patrones de muy grande. Esto a su vez causaba una saturación en las neuronas.**



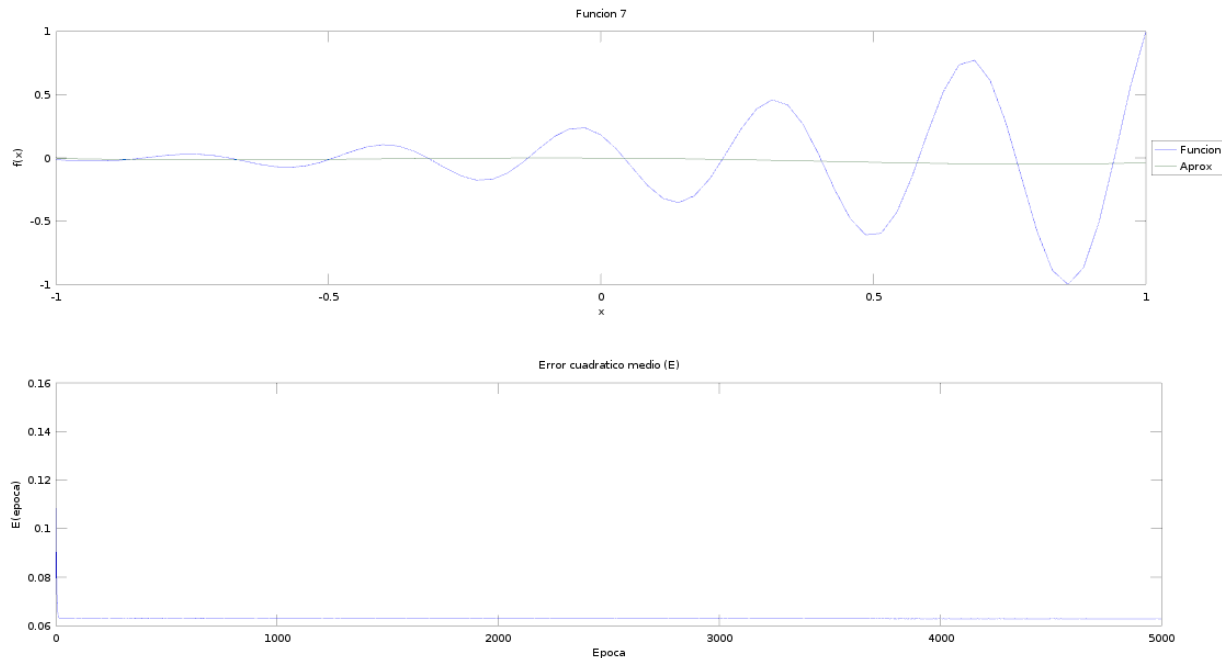
**¿Solución?**

# Normalizar

- Se llevó al intervalo  $[-1,1]$
- Al normalizar se trabajó con números más pequeños evitando la saturación de neuronas
- Trabajar en el intervalo  $[-1,1]$  funciona mejor con la función tangencial

**Entonces...**

# Prueba



$$\eta = 0.001$$

$$\beta = 1$$

arq = 1-35-10-1

g = tanh

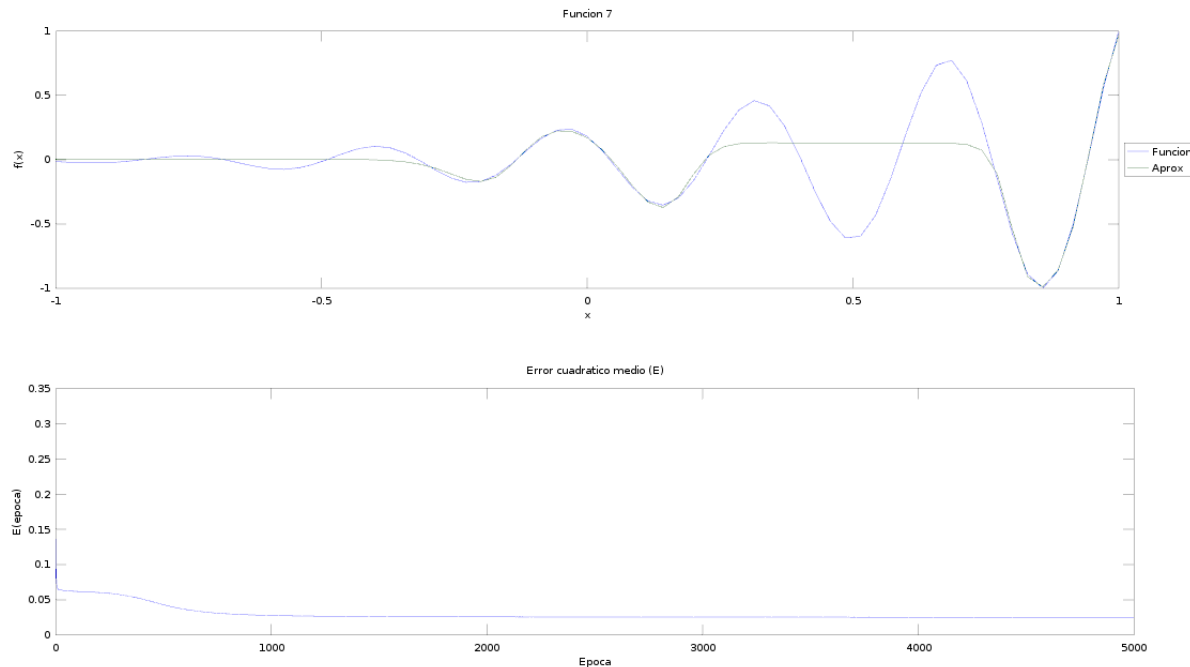
g salida = lineal

Épocas = 5000

**¿No funcionó?**

**Veamos otra prueba...**

# Prueba



$$\eta = 0.001$$

$$\beta = 3$$

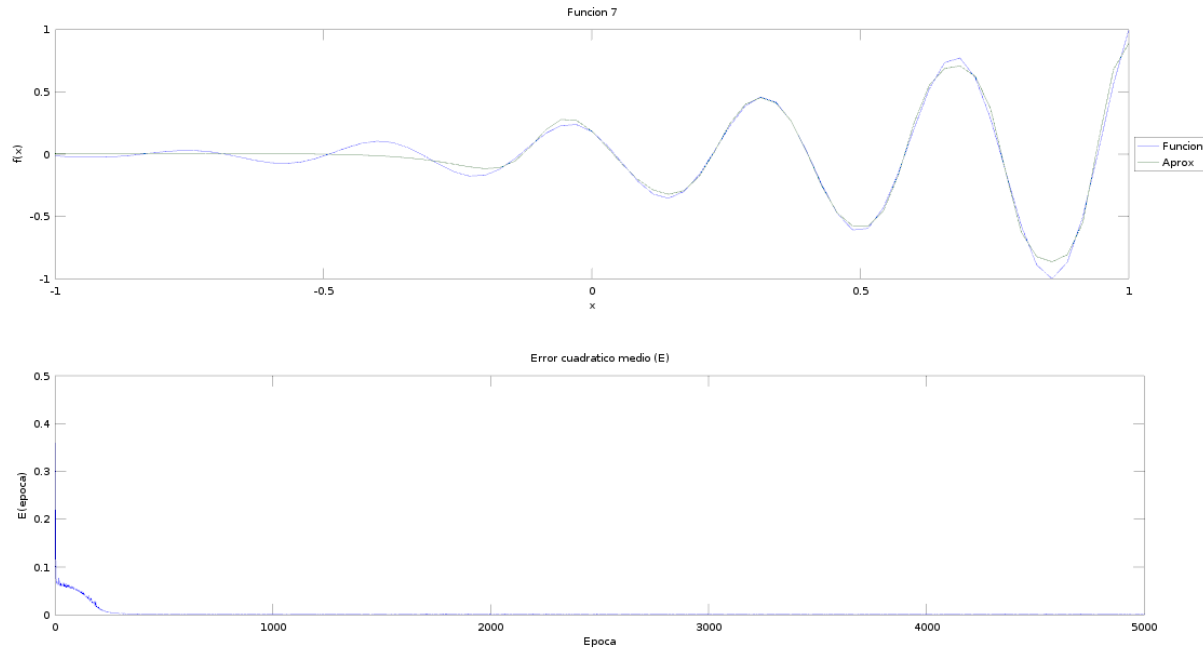
$$\text{arq} = 1-35-10-1$$

$$g = \tanh$$

$$g_{\text{salida}} = \text{lineal}$$

$$\text{Épocas} = 5000$$

# Otra prueba



$$\eta = 0.001$$

$$\beta = 3$$

$$\text{arq} = 1-35-10-1$$

$$g = \tanh$$

$$g \text{ salida} = \tanh$$

$$\text{Épocas} = 5000$$



# Prueba de diferentes arquitecturas

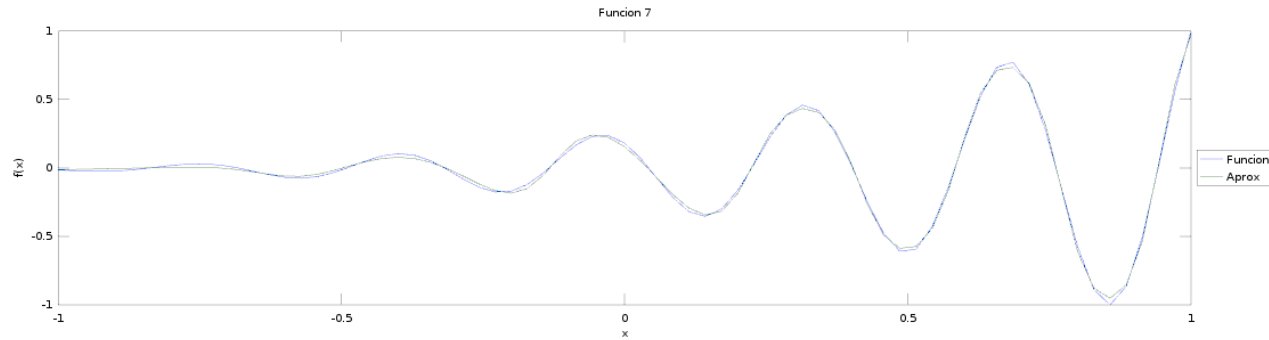
N	Arquitectura	$\beta$	$\eta$	$g(x)$	$g_{salida}(x)$	Épocas	$E(W)$
1	1-35-10-1	3	0,001	<i>tanh</i>	<i>lineal</i>	5000	0,025206
2	1-35-10-1	3	0,001	<i>tanh</i>	<i>tanh</i>	5000	0,001234
3	1-35-10-1	3	0,001	<i>exp</i>	<i>lineal</i>	5000	0,018987
4	1-35-10-1	3	0,001	<i>exp</i>	<i>tanh</i>	5000	0,008321
5	1-35-15-1	3	0,001	<i>tanh</i>	<i>lineal</i>	5000	0,000201
6	1-35-15-1	3	0,001	<i>tanh</i>	<i>tanh</i>	5000	0,000769
7	1-35-15-1	3	0,001	<i>exp</i>	<i>lineal</i>	5000	0,010085
8	1-35-15-1	3	0,001	<i>exp</i>	<i>tanh</i>	5000	0,001339
9	1-35-20-1	3	0,001	<i>tanh</i>	<i>lineal</i>	5000	0,000126
10	1-35-20-1	3	0,001	<i>tanh</i>	<i>tanh</i>	5000	0,000631
11	1-35-20-1	3	0,001	<i>exp</i>	<i>lineal</i>	5000	0,012053
12	1-35-20-1	3	0,001	<i>exp</i>	<i>tanh</i>	5000	0,00163
13	1-35-1	3	0,001	<i>tanh</i>	<i>tanh</i>	5000	0,059742

Tabla 1: Pruebas de las diferentes arquitecturas.

# Elección de la arquitectura

- Se eligió la arquitectura *1-35-15-1*
- $E(W)$  de un orden de magnitud menor que la arquitectura *1-35-10-1*
- En comparación con *1-35-20-1* en algunas pruebas ganaba una y en otras la otra pero la arquitectura elegida tenía una menor cantidad de neuronas por lo que requiere menos procesamiento
- Tener más neuronas en la segunda capa le permite distinguir mejor los patrones

# Arquitectura elegida



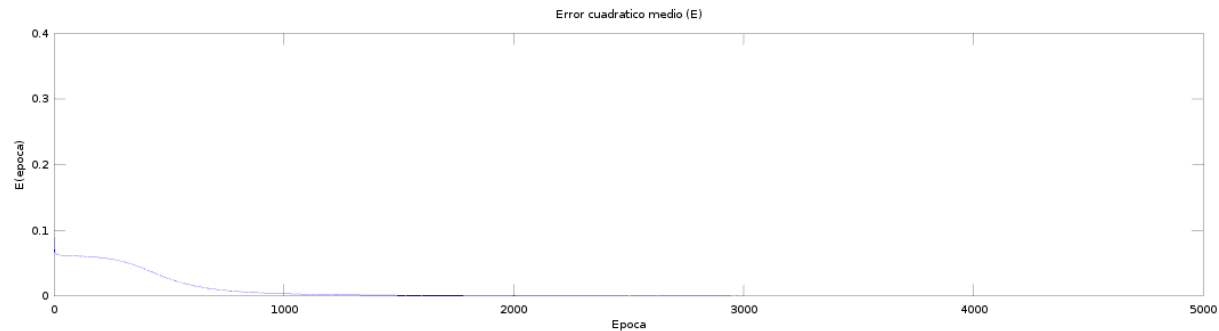
$$\eta = 0.001$$

$$\beta = 3$$

arq = 1-35-15-1

$g = \tanh$

$g_{\text{salida}} = \text{lineal}$

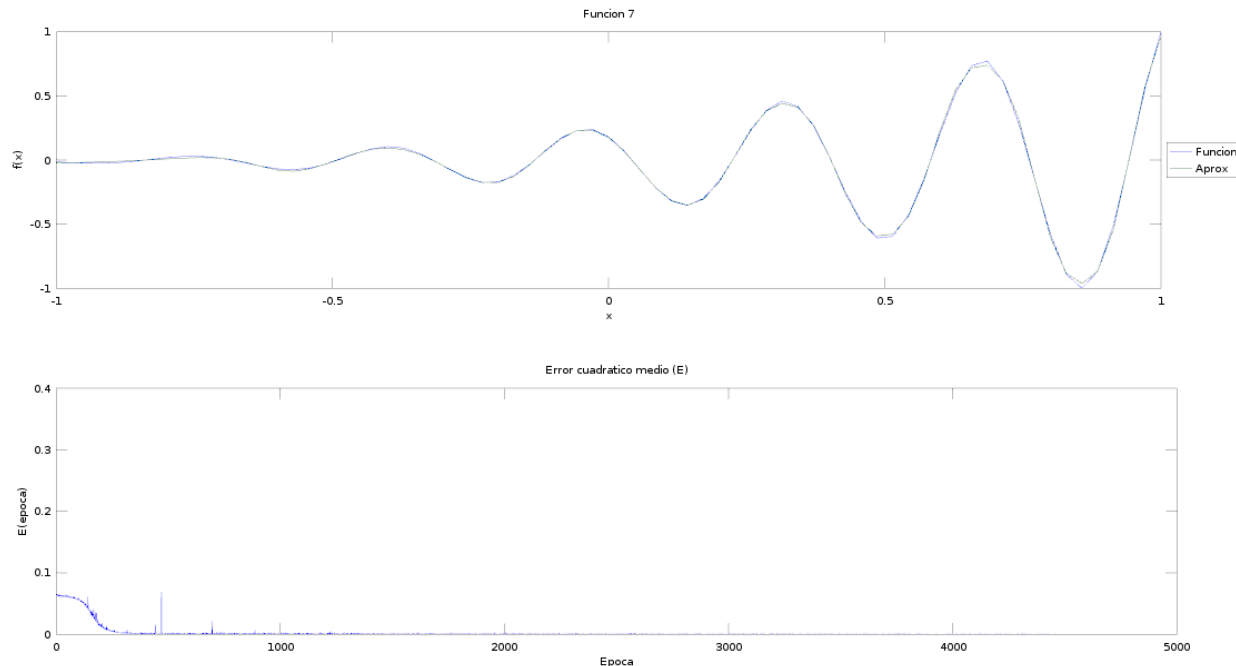


# Variaciones de $\eta$

N	Arquitectura	$\beta$	$\eta$	$g(x)$	$g_{salida}(x)$	Épocas	$E(W)$
1	1-35-15-1	3	0,005	$\tanh$	$lineal$	4467	0,000093
2	1-35-15-1	3	0,001	$\tanh$	$lineal$	5000	0,000201
3	1-35-15-1	3	0,0005	$\tanh$	$lineal$	5000	0,000412

Tabla 2: Pruebas de diferentes valores de  $\eta$ .

# Mejor $\eta$



$$\eta = 0.005$$

$$\beta = 3$$

$$\text{arq} = 1-35-15-1$$

$$g = \tanh$$

$$g \text{ salida} = \text{lineal}$$

$$E(W) = 0.00093$$

$$\acute{E}pocas = 4467$$

# Variaciones de $\beta$

N	Arquitectura	$\beta$	$\eta$	$g(x)$	$g_{salida}(x)$	Épocas	$E(W)$
1	1,35,15,1	2	0,005	$\tanh$	$lineal$	5000	0,001108
2	1,35,15,1	3, 5	0,005	$\tanh$	$lineal$	4594	0,000099
3	1,35,15,1	4	0,005	$\tanh$	$lineal$	5000	0,0003

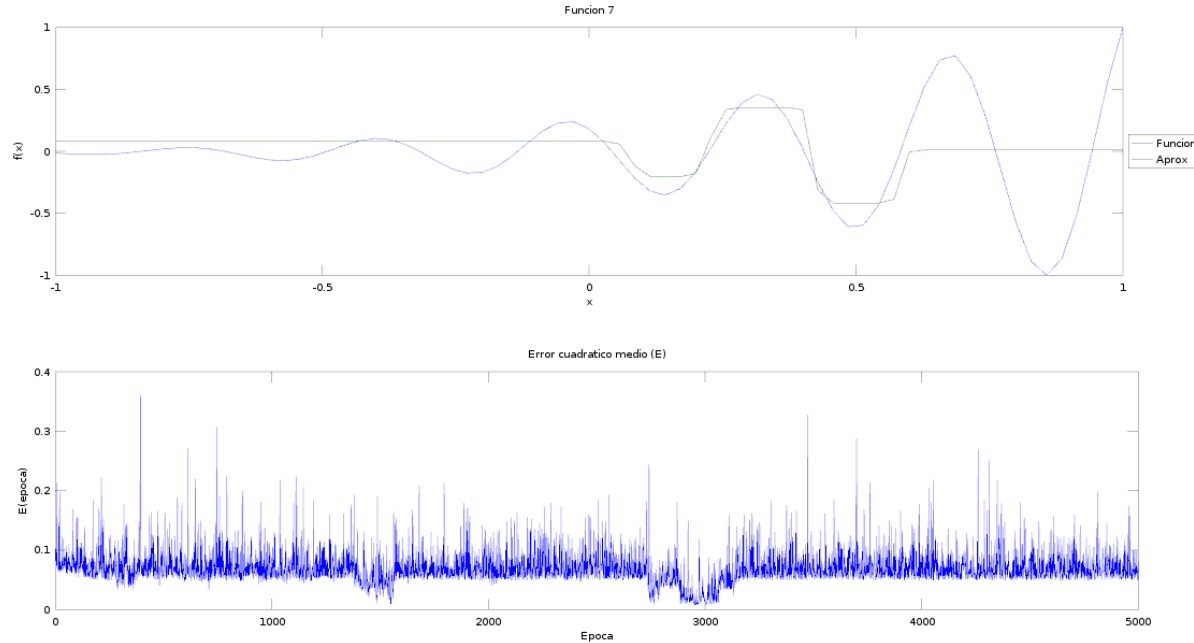
Tabla 3: Pruebas de diferentes valores de  $\beta$ .

# Momentum

N	Arquitectura	$\beta$	$\alpha$	$g(x)$	$g_{salida}(x)$	Épocas	$E(W)$
1	1-35-15-1	3	0,2	$\tanh$	$lineal$	3944	0,000096
2	1-35-15-1	3	0,3	$\tanh$	$lineal$	3700	0,000096
3	1-35-15-1	3	0,5	$\tanh$	$lineal$	5000	0.000306
4	1-35-15-1	3	0,7	$\tanh$	$lineal$	5000	0,000793
5	1-35-15-1	3	0,9	$\tanh$	$lineal$	5000	0,05111

Tabla 4: Pruebas de *momentum* variando  $\alpha$ .

# Peor momentum



$$\eta = 0.005$$

$$\beta = 3$$

$$\text{arq} = 1-35-15-1$$

$$g = \tanh$$

$$g \text{ salida} = \text{lineal}$$

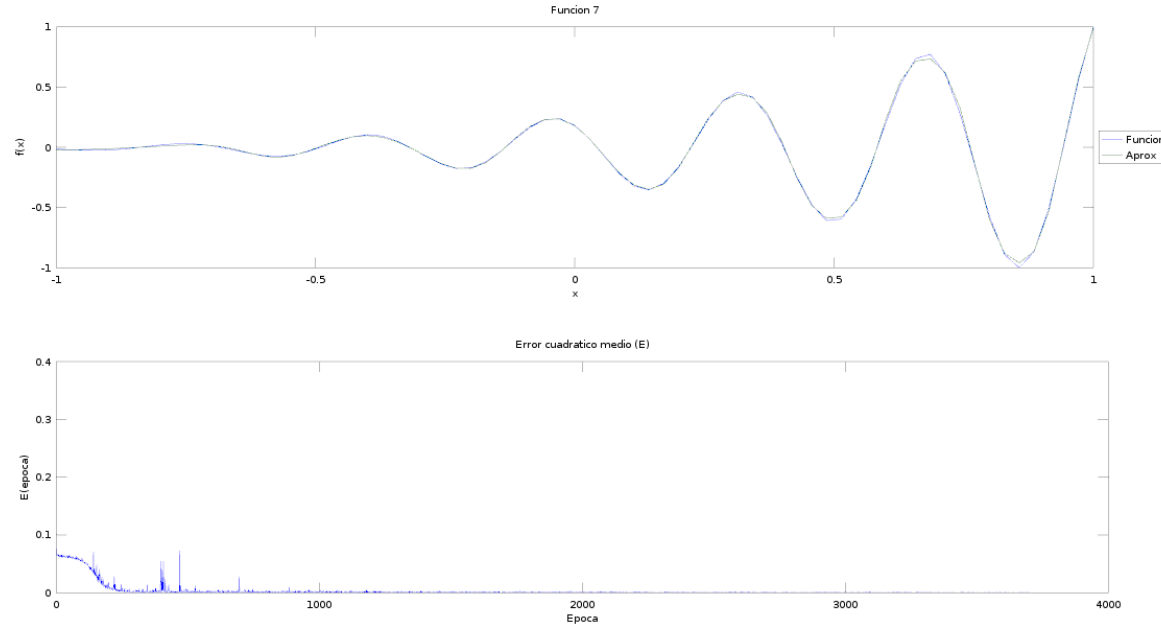
$$\alpha = 0.9$$

$$E(W) = 0.05111$$

En un *plateau*  $\eta \text{ efectivo} = \eta / (1 - \alpha)$



# Mejor momentum



$$\eta = 0.005$$

$$\beta = 3$$

$$\text{arq} = 1-35-10-1$$

$$g = \tanh$$

$$g \text{ salida} = \text{lineal}$$

$$\alpha = 0.9$$

$$E(W) = 0.00093$$

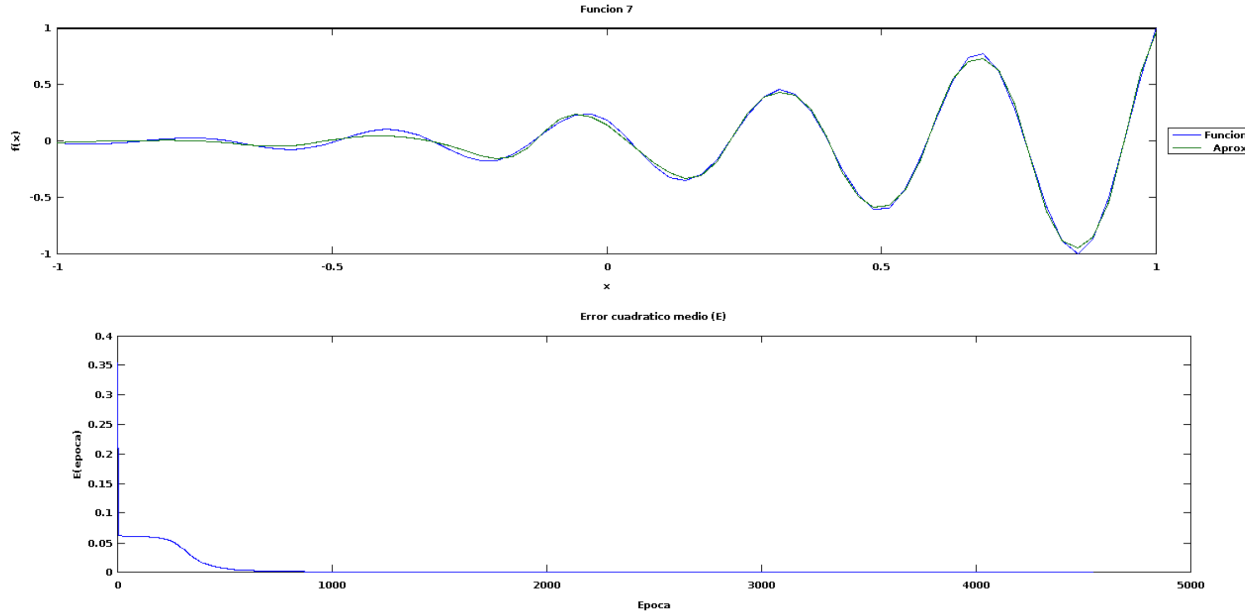
$$\text{Épocas} = 3700$$

# $\eta$ adaptativo

N	Arquitectura	$\beta$	$\eta_{final}$	$g(x)$	$g_{salida}(x)$	$a$	$b$	k	Épocas	$E(W)$
1	1,35,15,1	3	0,000242	$tanh$	$lineal$	0,0005	0,1	3	5000	0,000479
2	1,35,15,1	3	0,000329	$tanh$	$lineal$	0,0001	0,01	3	5000	0,00041
3	1,35,15,1	3	0,000311	$tanh$	$lineal$	0,001	0,05	3	5000	0,000422
4	1,35,15,1	3	0,000387	$tanh$	$lineal$	0,001	0,01	3	5000	0,000354
5	1,35,15,1	3	0,000333	$tanh$	$lineal$	0,0005	0,01	3	5000	0,000363

Tabla 5: Pruebas de  $\eta$  adaptativo variando los distintos parámetros.  $\eta_{final}$  es el valor de  $\eta$  al finalizar la prueba.

# Mejor $\eta$ adaptativo



$$\beta = 3$$

$$\text{arq} = 1-35-15-1$$

$$g = \tanh$$

$$g \text{ salida} = \text{lineal}$$

$$E(W) = 0.000354$$

$$a = 0.001$$

$$b = 0.01$$

$$k = 3$$

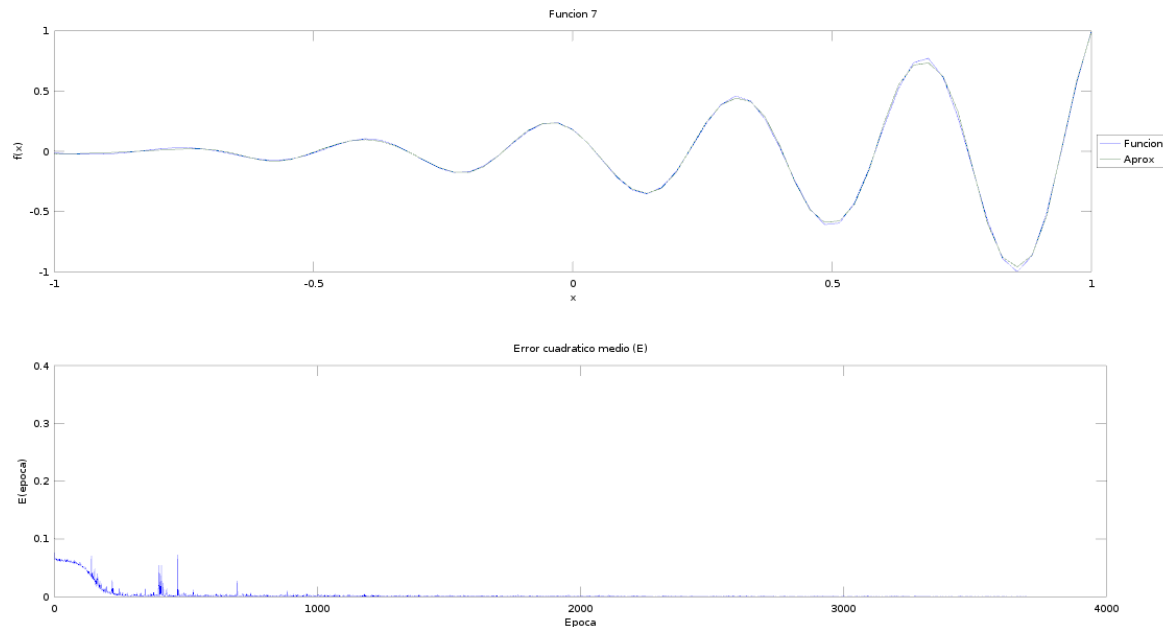
$$\text{Épocas} = 5000$$

# ***Momentum* + $\eta$ adaptativo**

N	Arquitectura	$\beta$	$g(x)$	$g_{salida}(x)$	$\alpha$	$a$	$b$	$k$	Épocas	$E(W)$
1	1,35,15,1	3	<i>tanh</i>	<i>lineal</i>	0,3	0,001	0,01	3	5000	0,000612
2	1,35,15,1	3	<i>tanh</i>	<i>lineal</i>	0,3	0,0005	0,01	3	5000	0,000592
3	1,35,15,1	3	<i>tanh</i>	<i>lineal</i>	0,3	0,0001	0,01	3	5000	0,00064

Tabla 6: Pruebas de  $\eta$  adaptativo con *momentum*.

# En resumen, la mejor combinación fue:



$$\eta = 0.005$$

$$\beta = 3$$

$$\text{arq} = 1-35-10-1$$

$$g = \tanh$$

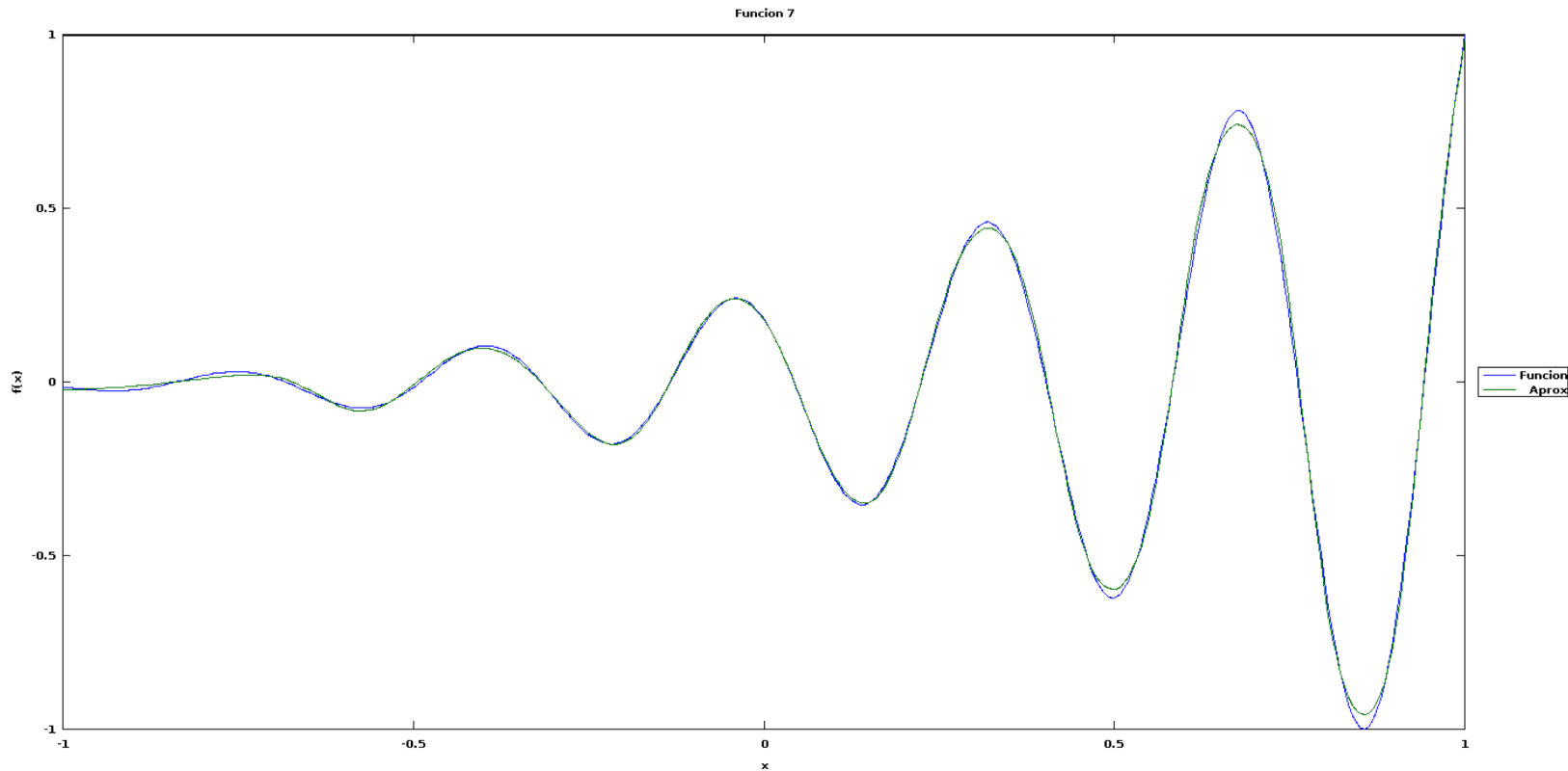
$$g \text{ salida} = \text{lineal}$$

$$\alpha = 0.9$$

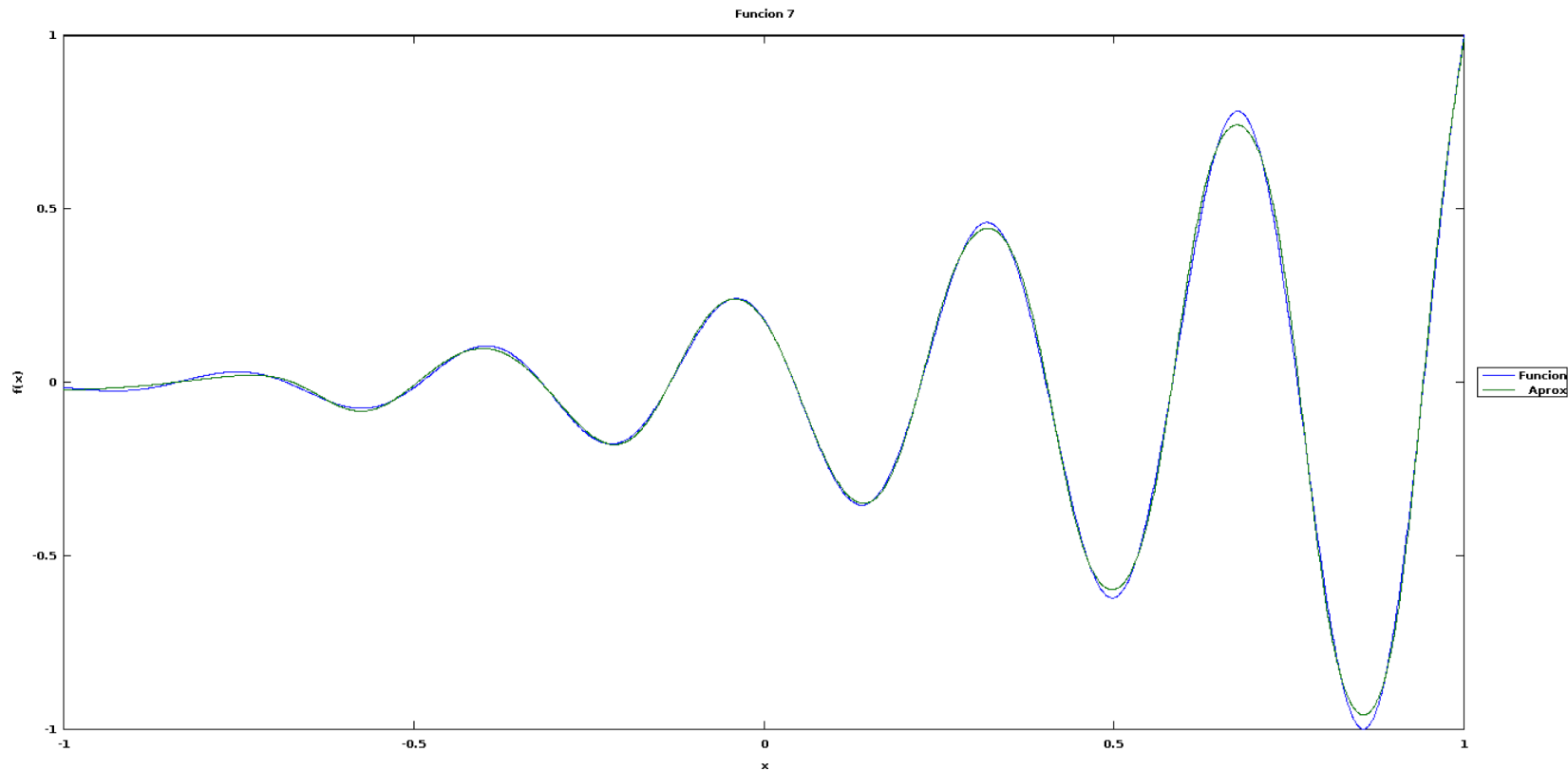
$$E(W) = 0.00093$$

$$\text{Épocas} = 3700$$

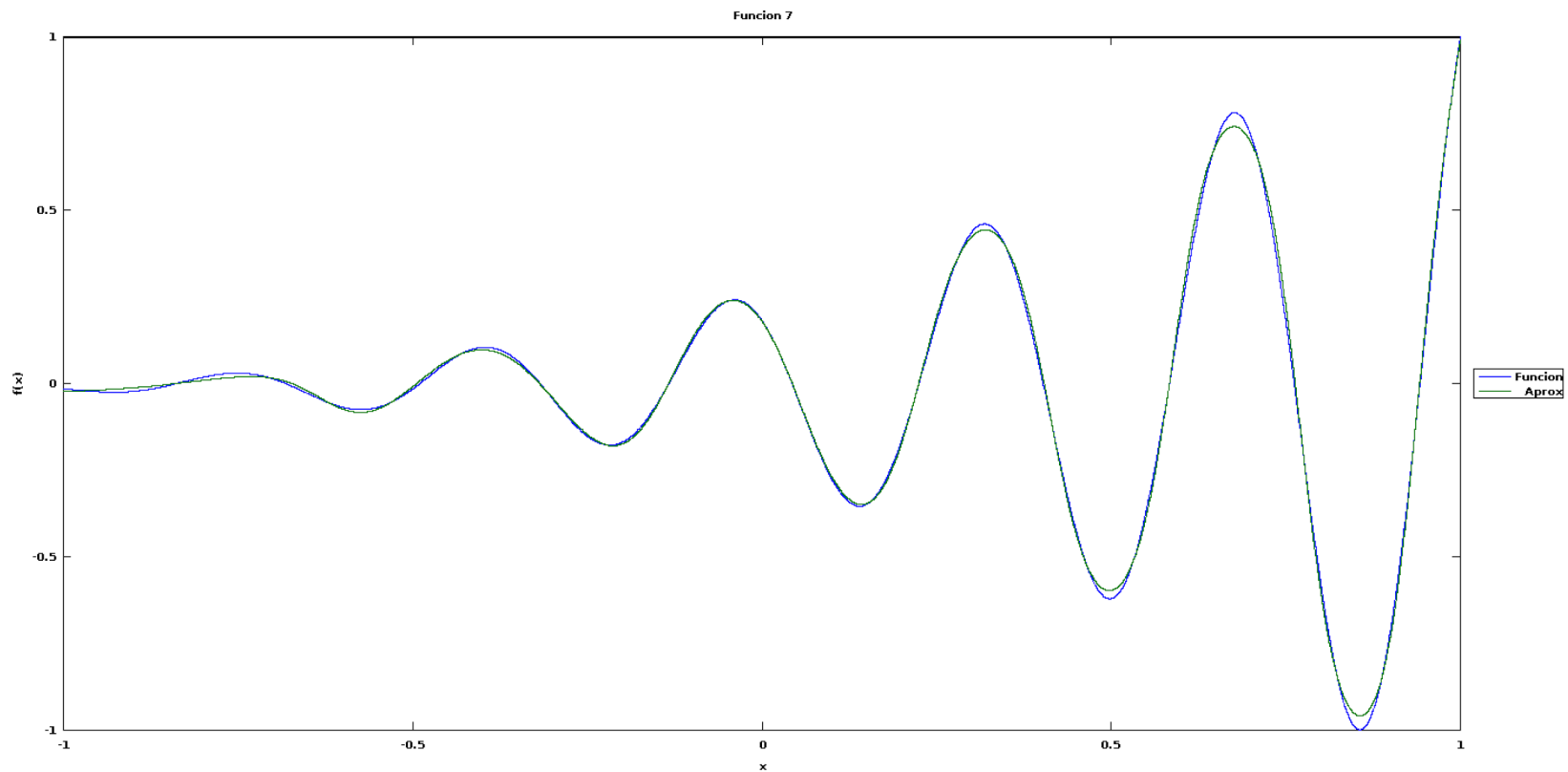
# Generalización (inc = 0.1)



# Generalización (inc = 0.01)



# Generalización (inc = 0.001)





# Conclusiones

- Se logró obtener una buena aproximación y una buena generalización del problema planteado
- Hay problemas que no se pueden solucionar con una única capa oculta
- La representación interna de los patrones es muy importante. Puede cambiar la dificultad del problema
- Si los patrones son muy grandes, normalizar es muy efectivo
- *Momentum* resultó ser la mejor mejora y superior a  $\eta$  adaptativo
- Más neuronas  $\neq$  mejor